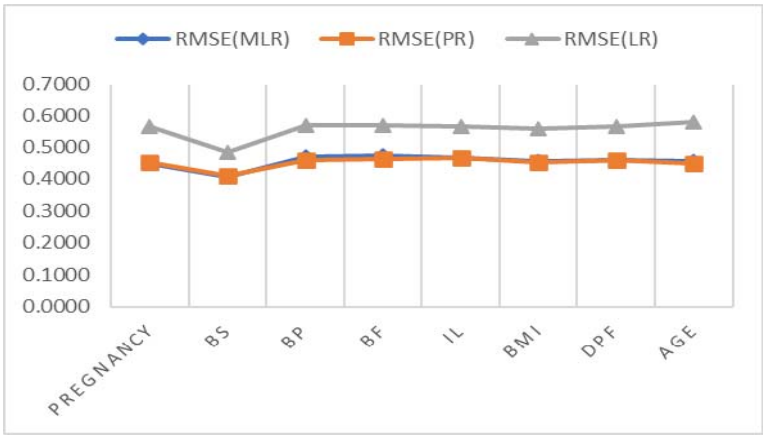
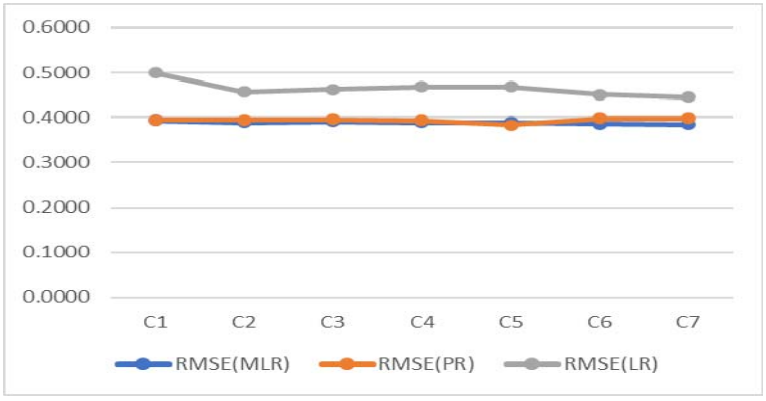


a) RMSE of PMS



b) RMSE of CPMS

Fig. 6. a) RMSE of PMS b) RMSE of CPMS

Accuracy of models is identified for each parameter as shown in Fig. 7. The entire PMS are computed for MLR, PR and LR models. R2 Score [37] identifies the accuracy of the models which is given in equation (4). Seeing the result, it is revealed that the best parameters for identifying the diabetes in an ordered arrangement are BS, Pregnancy, BMI, Age, DPF, BF, IL and BP.

$$R^2=1- \text{ RSS/TSS} \qquad (4)$$

where, RSS is Sum of squared residuals and TSS is Total sum of squares.

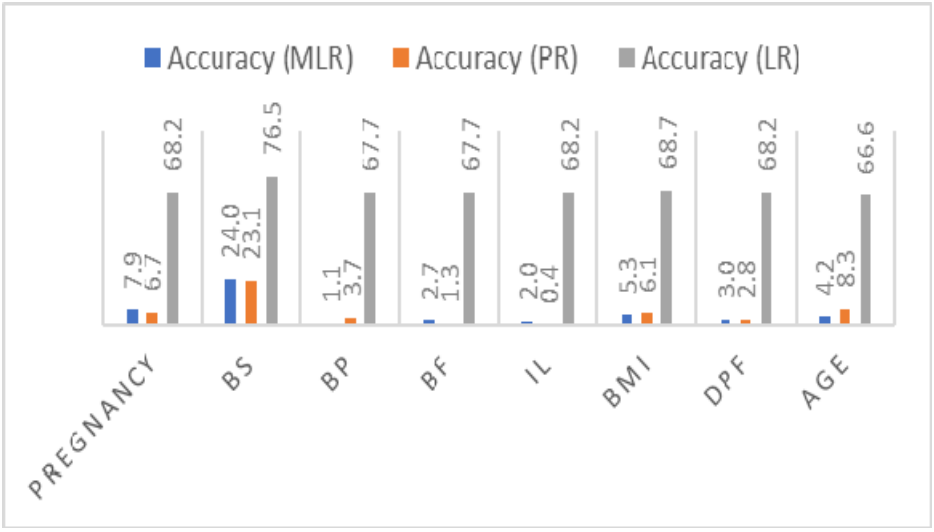


Fig. 7. Accuracy of PMS

A comparative analysis of MLR, PR and LR for each parameter in PMS is shown in Table 4. A significant percentage increase of accuracy is identified when Logistic Regression is used instead of Multiple Linear Regression and Polynomial Regression. The discrete nature of Logistic Regression is making a significant contribution in the enhancement of accuracy.

Table 4. Comparative Analysis of MLR, PR and LR for individual parameters

PMS	Accuracy (MLR)	Accuracy (PR)	Accuracy (LR)	Diff.(LR-MLR)	Diff.(LR-PR)	%Inc wrt MLR
Pregnancy	7.9	6.7	68.2	60.3	61.5	763.5
BS	24.0	23.1	76.5	52.5	53.4	218.8
BP	1.1	3.7	67.7	66.6	64.0	6054.5
BF	2.7	1.3	67.7	65.0	66.4	2407.4
IL	2.0	0.4	68.2	66.2	67.8	3310.0
BMI	5.3	6.1	68.7	63.4	62.6	1196.2
DPF	3.0	2.8	68.2	65.2	65.4	2173.3
Age	4.2	8.3	66.6	62.4	58.3	1485.7

Accuracy of the model is identified w.r.t. the ordered arrangement of parameters in Fig. 8. The parameters are combined together stepwise, so as to see if the accuracy of the model is enhanced or not. The entire CPMS are computed for the MLR, PR and Logistic Regression. R2 score is used to identify the accuracy of the model. Seeing the result, it is revealed that on uniting the parameters the accuracy is enhanced.

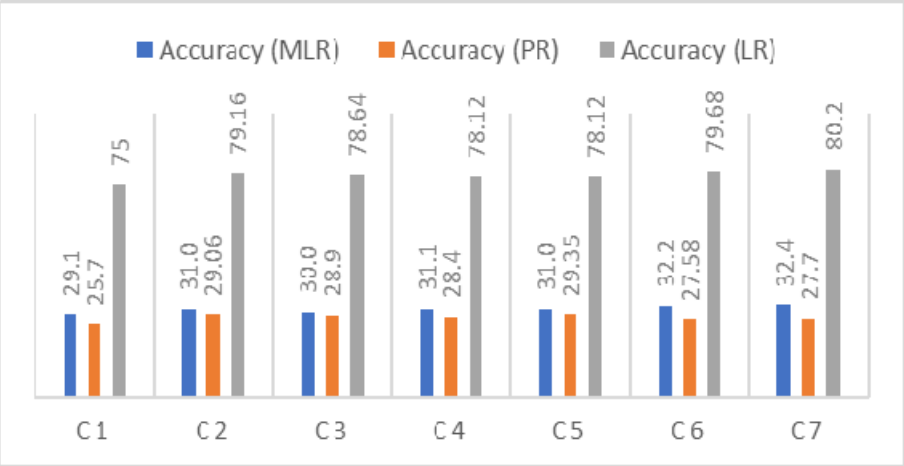


Fig. 8. Accuracy of CPMS for ordered combination of parameters

A comparative analysis of MLR, PR and LR for ordered combination of parameters in CPMS is shown in Table 5. A significant percentage increase of accuracy is identified when Logistic Regression is used instead of Multiple Linear Regression and Polynomial Regression.

Table 5. Comparative Analysis of MLR, PR and LR for ordered CPMS

CPMS	Accuracy (MLR)	Accuracy (PR)	Accuracy (LR)	Diff.(LR-MLR)	Diff.(LR-PR)	%Inc wrt MLR
C1	29.1	25.7	75	45.89	49.3	157.6
C2	31.0	29.06	79.16	48.13	50.1	155.1
C3	30.0	28.9	78.64	48.69	49.74	162.6
C4	31.1	28.4	78.12	47.06	49.72	151.5
C5	31.0	29.35	78.12	47.15	48.77	152.2
C6	32.2	27.58	79.68	47.48	52.1	147.5
C7	32.4	27.7	80.2	47.81	52.5	147.6

5. Conclusion

Diabetes stands as one of the most pervasive ailments afflicting people across the globe. This complex condition manifests through a range of parameters, often entailing severe and even life-threatening health implications. Swift identification of these parameters at an early stage holds the potential to preclude the emergence of severe health issues, offering substantial benefits to individuals. The current research is dedicated to a comprehensive parametric analysis of Patient Medical Statistics (PMS) and Combination of Patients Medical Statistics (CPMS) through the lens of linear, multi-linear, polynomial, and logistic regression models. The efficacy of these models is assessed using RMSE and R2 metrics, unveiling the logistic regression model's superior accuracy, attributed to its inherently discrete nature. Additionally, the study reveals that orchestrating CPMS in a systematic sequence according to their impact on diabetes detection yields heightened accuracy rates. Looking ahead, the study's implications suggest avenues for future exploration, such as broadening the scope of analyzed parameters for disease detection enhancement

References

[1] K. Gopinath, R. Jayakumararaj and M. Karthikeyan, "DAPD: A Knowledgebase for Diabetes Associated Proteins," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 12, no. 3, pp. 604-610, 1 May-June 2015, doi: 10.1109/TCBB.2014.2359442.

[2] C. Owens, H. Zisser, L. Jovanovic, B. Srinivasan, D. Bonvin and F. J. Doyle, "Run-to-run control of blood glucose concentrations for people with type 1 diabetes mellitus," in IEEE Transactions on Biomedical Engineering, vol. 53, no. 6, pp. 996-1005, June 2006, doi: 10.1109/TBME.2006.872818.

[3] M. E. Wilinska, L. J. Chassin, H. C. Schaller, L. Schaupp, T. R. Pieber and R. Hovorka, "Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin," in IEEE Transactions on Biomedical Engineering, vol. 52, no. 1, pp. 3-12, Jan. 2005, doi: 10.1109/TBME.2004.839639.

[4] S. Rahaman, "Diabetes diagnosis decision support system based on symptoms, signs and risk factor using special computational algorithm by rule base," 2012 15th International Conference on Computer and Information Technology (ICCIT), Chittagong, Bangladesh, 2012, pp. 65-71, doi: 10.1109/ICCITechn.2012.6509796.

[5] E. I. Georga et al., "Multivariate Prediction of Subcutaneous Glucose Concentration in Type 1 Diabetes Patients Based on Support Vector Regression," in IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 1, pp. 71-81, Jan. 2013, doi: 10.1109/TITB.2012.2219876.

[6] C. Owens, H. Zisser, L. Jovanovic, B. Srinivasan, D. Bonvin and F. J. Doyle, "Run-to-run control of blood glucose concentrations for people with type 1 diabetes mellitus," in IEEE Transactions on Biomedical Engineering, vol. 53, no. 6, pp. 996-1005, June 2006, doi: 10.1109/TBME.2006.872818.

[7] B. J. Lee and J. Y. Kim, "Identification of Type 2 Diabetes Risk Factors Using Phenotypes Consisting of Anthropometry and Triglycerides based on Machine Learning," in IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 1, pp. 39-46, Jan. 2016, doi: 10.1109/JBHI.2015.2396520.

[8] M. Cracchiolo et al., "Decoding Neural Metabolic Markers From the Carotid Sinus Nerve in a Type 2 Diabetes Model," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 10, pp. 2034-2043, Oct. 2019, doi: 10.1109/TNSRE.2019.2942398.

[9] P. Colmegna et al., "Evaluation of a Web-Based Simulation Tool for Self-Management Support in Type 1 Diabetes: A Pilot Study," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 1, pp. 515-525, Jan. 2023, doi: 10.1109/JBHI.2022.3209090.

[10] H. G. Clausen et al., "A New Stochastic Approach for Modeling Glycemic Disturbances in Type 2 Diabetes," in IEEE Transactions on Biomedical Engineering, vol. 68, no. 10, pp. 3161-3172, Oct. 2021, doi: 10.1109/TBME.2021.3074868.

[11] https://www.who.int/news-room/fact-sheets/detail/diabetes, [Date of Access: 06/04/2023]

[12] Mehan V., "Exploring the Future Jobs, Working Experience, Ethical Issues and Skills from Artificial Intelligence", International Journal of Innovative Science and Research Technology, Volume 8 , Issue 9 , Sept. - 2023.

[13] S. D. Patek et al., "Modular Closed-Loop Control of Diabetes," in IEEE Transactions on Biomedical Engineering, vol. 59, no. 11, pp. 2986-2999, Nov. 2012, doi: 10.1109/TBME.2012.2192930.

[14] W. Luangruangrong, A. Rodtook and S. Chimmanee, "Study of Type 2 diabetes risk factors using neural network for Thai people and tuning neural network parameters," 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Korea (South), 2012, pp. 991-996, doi: 10.1109/ICSMC.2012.6377858.

[15] C. Zhao and C. Yu, "Rapid Model Identification for Online Subcutaneous Glucose Concentration Prediction for New Subjects With Type I Diabetes," in IEEE Transactions on Biomedical Engineering, vol. 62, no. 5, pp. 1333-1344, May 2015, doi: 10.1109/TBME.2014.2387293.

[16] H. Abbas, L. Alic, M. Rios, M. Abdul-Ghani and K. Qaraqe, "Predicting Diabetes in Healthy Population through Machine Learning," 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Cordoba, Spain, 2019, pp. 567-570, doi: 10.1109/CBMS.2019.00117.

[17] X. Zhang, J. Deng and R. Su, "The EM algorithm for a linear regression model with application to a diabetes data," 2016 International Conference on Progress in Informatics and Computing (PIC), Shanghai, China, 2016, pp. 114-118, doi: 10.1109/PIC.2016.7949477.

[18] Y. Zou, X. Gong, P. Miao and Y. Liu, "Using TensorFlow to Establish multivariable linear regression model to Predict Gestational Diabetes," 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 2020, pp. 1695-1698, doi: 10.1109/ITNEC48623.2020.9084664.

[19] K. U. V. R. Teja, B. P. V. Reddy, L. P. A, H. Y. Patil and P. C. T, "Prediction of Diabetes at Early Stage with Supplementary Polynomial Features," 2021 Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 2021, pp. 1-5, doi: 10.1109/STCR51658.2021.9588849.

[20] A. Venkatesh. and M. S. Saravanan., "An Efficient Method for Predicting Linear Regression with Polynomial Regression," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1603-1606, doi: 10.1109/ICOSEC54921.2022.9952049.

[21] G. Oraz and X. Luo, "County-level geographic distributions of diabetes in relation to multiple factors in the united states," 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Las Vegas, NV, USA, 2018, pp. 279-282, doi: 10.1109/BHI.2018.8333423.

[22] F. Simone, F. Andrea, S. Giovanni, P. Gianluigi and D. F. Simone, "Linear Model Identification for Personalized Prediction and Control in Diabetes," in IEEE Transactions on Biomedical Engineering, vol. 69, no. 2, pp. 558-568, Feb. 2022, doi: 10.1109/TBME.2021.3101589.
[23] V. Ganesh, J. Kolluri and K. V. Kumar, "Diabetes Prediction using Logistic Regression and Feature Normalization," 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2021, pp. 1-6, doi: 10.1109/ICSES52305.2021.9633773.
[24] L. Lei, "Prediction of Score of Diabetes Progression Index Based on Logistic Regression Algorithm," 2020 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), Zhangjiajie, China, 2020, pp. 954-956, doi: 10.1109/ICVRIS51417.2020.00232.
[25] A. Mangal and V. Jain, "Performance analysis of machine learning models for prediction of diabetes," 2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 2022, pp. 1-4, doi: 10.1109/CISCT55310.2022.10046630.
[26] M. F. Faruque, Asaduzzaman and I. H. Sarker, "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-4, doi: 10.1109/ECACE.2019.8679365.
[27] T. H. Nasution and L. A. Harahap, "Predict the Percentage Error of LM35 Temperature Sensor Readings using Simple Linear Regression Analysis," 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), Medan, Indonesia, 2020, pp. 242-245, doi: 10.1109/ELTICOM50775.2020.9230472.
[28] C. -H. Wu, J. -B. Li and T. -Y. Chang, "SLinRA2S: A Simple Linear Regression Analysis Assisting System," 2013 IEEE 10th International Conference on e-Business Engineering, Coventry, UK, 2013, pp. 219-223, doi: 10.1109/ICEBE.2013.33.
[29] P. Wang, R. Ge, X. Xiao, M. Zhou and F. Zhou, "hMuLab: A Biomedical Hybrid MUlti-LABel Classifier Based on Multiple Linear Regression," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 14, no. 5, pp. 1173-1180, 1 Sept.-Oct. 2017, doi: 10.1109/TCBB.2016.2603507.
[30] H. T. Hoc, R. Silhavy, Z. Prokopova and P. Silhavy, "Comparing Multiple Linear Regression, Deep Learning and Multiple Perceptron for Functional Points Estimation," in IEEE Access, vol. 10, pp. 112187-112198, 2022, doi: 10.1109/ACCESS.2022.3215987.
[31] Y. Wang, L. Li and C. Dang, "Calibrating Classification Probabilities with Shape-Restricted Polynomial Regression," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 8, pp. 1813-1827, 1 Aug. 2019, doi: 10.1109/TPAMI.2019.2895794.
[32] A. Venkatesh. and M. S. Saravanan., "An Efficient Method for Predicting Linear Regression with Polynomial Regression," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1603-1606, doi: 10.1109/ICOSEC54921.2022.9952049.
[33] K. He and C. He, "Housing Price Analysis Using Linear Regression and Logistic Regression: A Comprehensive Explanation Using Melbourne Real Estate Data," 2021 IEEE International Conference on Computing (ICOCO), Kuala Lumpur, Malaysia, 2021, pp. 241-246, doi: 10.1109/ICOCO53166.2021.9673533.
[34] J. Nie, J. Fang and Y. Zhao, "Cow Health Prediction Method Based on Logistic Regression and Decision Tree," 2022 34th Chinese Control and Decision Conference (CCDC), Hefei, China, 2022, pp. 3712-3717, doi: 10.1109/CCDC55256.2022.10033946.
[35] https://www.kaggle.com/datasets/mathchi/diabetes-data-set, Date of Access: 28-04-2023
[36] K. Rajesh and M. S. Saravanan, "Prediction of Customer Spending Score for the Shopping Mall using Gaussian Mixture Model comparing with Linear Spline Regression Algorithm to reduce Root Mean Square Error," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 335-341, doi: 10.1109/ICICCS53718.2022.9788162.
[37] Gregory Y.H. Lip, Ken Haguenoer, Christophe Saint-Etienne, Laurent Fauchier, "Relationship of the SAMe-TT2 R2 Score to Poor-Quality Anticoagulation, Stroke, Clinically Relevant Bleeding, and Mortality in Patients With Atrial Fibrillation," Chest, Volume 146, Issue 3, 2014, Pages 719-726, ISSN 0012-3692, https://doi.org/10.1378/chest.13-2976.

Conflict of Interest: The author has no conflicts of interest to declare.

Aut**hor Profile**

Prof. (Dr.) Vineet Mehan, received the B.Tech. degree from Kurukshetra University, M.E. degree from NITTTR Chandigarh and Ph.D. degree from NIT Jalandhar. He is currently a Professor with the Department of Artificial Intelligence and Machine Learning, Computer Science and Engineering, Chandigarh University, Punjab, India, since Jun. 2023. He has published over 50 papers in peer-reviewed international journals and conferences. Her research interests include Machine Learning, AI, Bio-informatics and DIP..